

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:24 ; Search time 91.75 Seconds

(without alignments)  
15.839 Million cell updates/sec

Title: us-09-331-631a-1\_copy\_29\_73

Perfect score: 252  
Sequence: 1 SEEDROETECKRQCMQLET.....RCVSQCKRFEEDIMSKYD 45

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.5	30.8	588	1 VCLB_GOSHI	P09801 gossypium h
2	72	28.6	605	1 VCLB_GOSHI	P09799 gossypium h
3	67.5	26.8	47	1 AGRP_LURCY	P36368 lufta cylin
4	67	26.6	1170	1 TSP1_HUMAN	P07996 homo sapien
5	67	26.6	1170	1 TSP1_MOUSE	P35441 mus musculu
6	65	25.8	1173	1 TSP1_XENLA	P35448 xenopus lae
7	63	25.0	524	1 SBP_SOYBN	P04672 glycine max
8	57	22.6	724	1 ATIL_VACCV	P24759 vaccinia vi
9	57	22.6	726	1 ATIL_CAMPC	P05482 camelpox vi
10	57	22.6	1284	1 ATIL_CAMPC	P05482 camelpox vi
11	55	21.8	305	1 PTP6_RAT	P16602 compox viru
12	55	21.8	795	1 ENPL_CHICK	P08110 gallus norv
13	54.5	21.6	244	1 YL53_CAREL	P44433 caenorhabdi
14	54.5	21.6	1021	1 YL53_CAREL	P46582 caenorhabdi
15	54	21.4	291	1 CGD1_BRARE	P09459 brachydanio
16	54	21.4	301	1 RO60_CAREL	P07274 caenorhabdi
17	53.5	21.2	301	1 G197_HUMAN	P04457 homo sapien
18	53.5	21.2	1170	1 TSP2_BOVIN	P09516 bos taurus
19	53	21.0	702	1 ATIL_VARY	P34011 variola vir
20	52.5	20.8	91	1 UCRH_HUMAN	P07919 homo sapien
21	52.5	20.8	1172	1 TSP2_HUMAN	P35442 homo sapien
22	52	20.6	33	1 MBP1_MAIZE	P28794 zea mays (m
23	52	20.6	37	1 CG2S_LUPAN	P09930 lupinus ang
24	52	20.6	82	1 C2_OXYNO	P05526 oxytricha n
25	52	20.6	716	1 ENPL_RABIT	P08113 mus musculu
26	52	20.6	803	1 ENPL_MOUSE	P14148 homo sapien
27	52	20.6	803	1 ENPL_MOUSE	P14148 homo sapien
28	52	20.6	804	1 TIR_ECOLI	P08956 escherichia
29	52	20.6	1188	1 COLA_HORSE	P02704 equus cabal
30	51.5	20.4	106	1 COLA_HORSE	P02705 equus cabal
31	51.5	20.4	108	1 COLB_HORSE	P02705 equus cabal
32	51.5	20.4	147	1 LAPP_HAEOF	P01747 haementeria
33	51.5	20.4	298	1 HIST_CANAL	P46386 candida alb

34	51.5	20.4	525	1 Y107_METJA	O57571 methanococ
35	51.5	20.4	563	1 MUCCS_BOVIN	P98091 bos taurus
36	51.5	20.4	860	1 UL52_HSV6U	P52467 herpes simp
37	51.5	20.4	860	1 UL52_HSV6Z	P52467 herpes simp
38	51.5	20.4	1557	1 DVAL_DICVI	O24702 dictyocaulu
39	51	20.2	418	1 CDL5_HUMAN	Q14004 homo sapien
40	51	20.2	741	1 FIBA_CHICK	P14448 gallus gall
41	51	20.2	881	1 ARP8_YEAST	Q12366 saccharomyc
42	51	20.2	893	1 YM92_CAREL	P34531 caenorhabdi
43	51	20.2	1231	1 CEAH_HUMAN	P08603 homo sapien
44	51	20.2	1370	1 Z261_HUMAN	O14202 homo sapien
45	50.5	20.0	264	1 RPOD_METTH	O26144 methanodact

## ALIGNMENTS

RESULT 1					
VCLB_GOSHI		STANDARD;	PRT;	588 AA.	
ID VCLB_GOSHI					
AC P09801;					
DT 01-MAR-1989 (Rel. 10, Created)					
DT 01-MAR-1989 (Rel. 10, Last sequence update)					
DT 15-JUL-1999 (Rel. 38, Last annotation update)					
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).					
OS Gossypium hirsutum (Upland cotton).					
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;					
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;					
OC Malvales; Malvaceae; Gossypium.					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;					
RT "Developmental biochemistry of cottonseed embryogenesis and					
RT germination. XVIII. cDNA and amino acid sequences of the members of					
RT the storage protein families."					
RL Plant Mol. Biol. 7:475-489(1986).					
CC -I- FUNCTION: SEED STORAGE PROTEIN.					
CC -I- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN					
CC BODIES.					
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASOLIN, VICILIN,					
CC CONVICTILIN, CONGLYCININ, ETC.).					
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration					
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC					
DR EMBL; M16891; AAA33071.1; -.					
DR PIR; A30838; FMCNAB.					
DR HSSP; P50477; ICAX.					
DR INTERPRO; IPR001113; -.					
DR PFAM; PF00546; Seedstore_7s; 1.					
KW Seed storage protein; Signal.					
FT SIGNAL 1					
FT CHAIN 26					
FT SEQUENCE 588 AA; 69729 MW; 63E699B29A8A8ADB CRC64;					
Query Match		30.8%;	Score 77.5;	DB 1;	Length 588;
Best Local Similarity		40.5%;	Pred. No. 0.059;		
Matches 15;		Conservative 9;	Mismatches 12;	Indels 1;	Gaps 1;
QY 2 EFDRQETECKRQCMQLETSGQMRVCVSCQDKRFEED 38					
Db 81 EDPQRYEECKRQCMQLETSGQMRVCVSCQDKRFEED 116					
RESULT 2					
VCLB_GOSHI		STANDARD;	PRT;	605 AA.	
ID VCLB_GOSHI					

AC P09799;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).  
OS Gossypium hirsutum (Upland cotton).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Malvales; Malvaceae; Gossypium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;  
RT "Developmental biochemistry of cottonseed embryogenesis and  
germination. XIX. Sequences and genomic organization of the alpha  
globulin (vicilin) genes of cottonseed.";  
RL Plant Mol. Biol. 9:533-546(1987).  
CC -1- FUNCTION: SEED STORAGE PROTEIN.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN  
BODIES.  
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
CONVULSIN, CONGLICININ, ETC.).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M19378; AAA33069.1; -;  
DR PIR: S06398; S06398.  
DR HSSP: P50477; 1CAX.  
DR INTERPRO: IPR001113; -;  
DR PFM: PF00546; Seedstore.7s; 1.  
KM Seed storage protein: Signal.  
FT SIGNAL 1 23  
FT CHAIN 24 605 VICILIN GC72-A.  
SQ SEQUENCE 605 AA; 71049 MM; C9DB9371C976953B CRC64;  
  
Query Match 28.8%; Score 72; DB 1; Length 605;  
Best Local Similarity 42.5%; Pred. No. 0.26; 10; Indels 4; Gaps 3;  
Matches 17; Conservative 9; Mismatches 10;  
  
QY 1 SEED-ROVEECKRQCMQLETSQMRVCVSCDKRFEEDI 39  
ID 33 SEDDPQQRVEDCKRRC-QLETRQGTED--DKCEDRSETQL 69  
DB 33 SEDDPQQRVEDCKRRC-QLETRQGTED--DKCEDRSETQL 69  
  
RESULT 3  
ACRP\_LUFCY STANDARD; PRT; 47 AA;  
ID AGRP\_LUFCY  
AC P56568;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 6.5 KDA ARGININE/GUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).  
OS Luifia cylindrica (Smooth loofah) (Sponge gourd).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Cucurbitales; Cucurbitaceae; Luifia.  
RN [1]  
RP SEQUENCE.  
RC TISSUE-SEED.  
RX MEDLINE=97357433; PubMed=9214759;  
RA Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;  
RT "Primary structure of 6.5k-arginine/glutamate-rich polypeptide from  
the seeds of sponge gourd (Luifia cylindrica).";  
RL Biosci. Biotechnol. Biochem. 61:984-988(1997).  
CC -1- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON  
RESERVES DURING GERMINATION AND SEEDLING GROWTH.  
CC -1- MASS SPECTROMETRY: MM-3693.39; METHOD-MALDI.

CC -1- SIMILARITY: SOME, TO 7S SEED STORAGE PROTEINS.  
KM Seed storage protein.  
FT DISULFID 12 33  
FT DISULFID 16 29  
SQ SEQUENCE 47 AA; 5698 MM; 588B0EC82273AC05 CRC64;  
  
Query Match 26.8%; Score 67.5; DB 1; Length 47;  
Best Local Similarity 41.2%; Pred. No. 0.081;  
Matches 14; Conservative 4; Mismatches 15; Indels 1; Gaps 1;  
  
QY 5 ROVEECKRQCMQLETS-QMRRVCVSCDKRFEEDI 37  
ID 6 RTEYACRVRQYAEHGVRRQRCQYCEKRLRE 39  
DB 6 RTEYACRVRQYAEHGVRRQRCQYCEKRLRE 39  
  
RESULT 4  
TSPL\_HUMAN STANDARD; PRT; 1170 AA.  
ID TSPL\_HUMAN  
AC P07996;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE THROMBOSPONDIN, I PRECURSOR.  
GN THBS1 OR TSPI OR TSP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-ENDOTHELIAL CELLS;  
RX MEDLINE=87057617; PubMed=2430973;  
RA Lawler J., Hynes R.O.;  
RT "The structure of human thrombospondin, an adhesive glycoprotein with  
multiple calcium-binding sites and homologies with several different  
proteins.";  
RL J. Cell Biol. 103:1635-1648(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89139590; PubMed=2918029;  
RA Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,  
BA Baumgartel D.M., Rotwein P., Frazier W.A.;  
RT "Complete thrombospondin mRNA sequence includes potential regulatory  
sites in the 3' untranslated region.";  
RL J. Cell Biol. 108:729-736(1989).  
RN [3]  
RP SEQUENCE OF 1-397 FROM N.A.  
RX MEDLINE=87157592; PubMed=3030396;  
RA Kobayashi S., Eden-McCutchan F., Framson P., Bornstein P.;  
RT "Partial amino acid sequence of human thrombospondin as determined by  
analysis of cDNA clones: homology to malarial circumsporozoite  
proteins.";  
RL Biochemistry 25:8418-8425(1986).  
RN [4]  
RP SEQUENCE OF 1-374 FROM N.A.  
RX MEDLINE=86287276; PubMed=3461443;  
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;  
RT "Characterization of a cDNA encoding the heparin and collagen binding  
domains of human thrombospondin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).  
RN [5]  
RP SEQUENCE OF 1-166 FROM N.A.  
RX MEDLINE=89291870; PubMed=2544587;  
RA Laherty C.D., Gierman T.M., Dixit V.M.;  
RT "Characterization of the promoter region of the human thrombospondin  
gene. DNA sequences within the first intron increase transcription.";  
RL J. Biol. Chem. 264:11222-11227(1989).  
RN [6]  
RP SEQUENCE OF 1028-1170 FROM N.A.  
RA la Fleur M., Jobin C., Gauthier J., Kreis C.G.;  
RT Submitted (XXX-1992) to the EMBL/Genbank/DDP databases.  
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,

[illegible]

```

FT CAROHWD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHWD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHWD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHWD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 84 84 T -> A (IN REF. 2, 3 AND 4).
FT CONFLICT 523 523 T -> A (IN REF. 2).
SQ SEQUENCE 1170 AA; 129412 MW; 69B3BDE5AE3A395E CRC64;

Query Match 26.6%; Score 67; DB 1; Length 1170;
Best Local Similarity 39.5%; Pred. No. 1.9;
Matches 15; Conservative 6; Mismatches 11; Indels 6; Gaps 2

OY 13 ROCMOLE-----TSGOMRRC-VSCDKRFEEDIDMSKY 44
      | | | | | | | | | | | | | | | | | | | | | |
DB 404 RSCDSLNRRCGSSVOTRTCHIOECDKPRKQGGMSHW 441

RESULT 5
TSPL_MOUSE STANDARD: PRT; 1170 AA.
AC TSPL_MOUSE P35441;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=92128941; PubMed=1774063;
RX Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
RX Jenkins N.A.;
RA "Characterization of the murine thrombospondin gene.";
RA Genomics 11:587-600(1991).
RL [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RX Laherty J., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
RT during cell growth and development.";
RL J. Biol. Chem. 267:3274-3281(1992).
RN [3]
RP SEQUENCE OF 1-490 FROM N.A.
RX MEDLINE=90375546; PubMed=2398070;
RX Bornstein P., Allt D., Devarayalu S., Framson P., Li P.;
RA "Characterization of the mouse thrombospondin gene and evaluation of
RA the role of the first intron in human gene expression.";
RL J. Biol. Chem. 265:16691-16698(1990).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VMRC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS
CC (WHICH BIND CALCIUM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
RR EMBL; M62470; AAA50611.1; -
RR EMBL; M62450; AAA50611.1; JOINED.
RR EMBL; M62451; AAA50611.1; JOINED.

```

DR	EMBL	M62452	AAA50611.1	JOINED.
DR	EMBL	M62453	AAA50611.1	JOINED.
DR	EMBL	M62454	AAA50611.1	JOINED.
DR	EMBL	M62455	AAA50611.1	JOINED.
DR	EMBL	M62456	AAA50611.1	JOINED.
DR	EMBL	M62457	AAA50611.1	JOINED.
DR	EMBL	M62458	AAA50611.1	JOINED.
DR	EMBL	M62459	AAA50611.1	JOINED.
DR	EMBL	M62460	AAA50611.1	JOINED.
DR	EMBL	M62461	AAA50611.1	JOINED.
DR	EMBL	M62462	AAA50611.1	JOINED.
DR	EMBL	M62463	AAA50611.1	JOINED.
DR	EMBL	M62464	AAA50611.1	JOINED.
DR	EMBL	M62465	AAA50611.1	JOINED.
DR	EMBL	M62466	AAA50611.1	JOINED.
DR	EMBL	M62467	AAA50611.1	JOINED.
DR	EMBL	M62468	AAA50611.1	JOINED.
DR	EMBL	M62469	AAA50611.1	JOINED.
DR	EMBL	M87276	AAA53063.1	-.
DR	EMBL	J05606	AAA40431.1	-.
DR	EMBL	J05605	AAA40431.1	JOINED.
DR	PIR	A40558	A40558.	
DR	PIR	B42587	B42587.	
DR	PIR	A37905	A37905.	
DR	HSSP	P33555	1EMO.	
DR	MGD	MGI:98737	THBS1.	
DR	INTERPRO	IPR000561	-.	
DR	INTERPRO	IPR000884	-.	
DR	INTERPRO	IPR001007	-.	
DR	PFAM	PF00008	EGF_2.	
DR	PFAM	PF00090	TSP_1; 3.	
DR	PFAM	PF00093	wvc_1.	
DR	PROSITE	PS00022	EGF_1; FALSE-NEG.	
DR	PROSITE	PS01186	EGF_2; 1.	
DR	PROSITE	PS50092	TSP1; 3.	
DR	PROSITE	PS01208	WVFC; 1.	
KW	Glycoprotein	Cell adhesion	Calcium-Binding; Heparin-Binding; Repeat	
KW	EGF-like domain	Signal	POTENTIAL.	
FT	CHAIN	1	1170	
FT	DOMAIN	19	232	
FT	DOMAIN	316	373	
FT	DOMAIN	379	548	
FT	DOMAIN	549	690	
FT	DOMAIN	723	950	
FT	DOMAIN	951	1170	
FT	REPEAT	379	430	
FT	REPEAT	435	491	
FT	REPEAT	492	548	
FT	DOMAIN	549	587	
FT	DOMAIN	588	645	
FT	DOMAIN	646	690	
FT	REPEAT	723	758	
FT	REPEAT	759	781	
FT	REPEAT	782	817	
FT	REPEAT	818	840	
FT	REPEAT	841	878	
FT	REPEAT	879	914	
FT	REPEAT	915	950	
FT	SITE	926	928	
FT	DISULFID	270	270	
FT	DISULFID	274	274	
FT	DISULFID	551	562	
FT	DISULFID	556	572	
FT	DISULFID	575	586	
FT	DISULFID	592	608	
FT	DISULFID	599	617	
FT	DISULFID	620	644	
FT	DISULFID	650	663	
FT	DISULFID	657	676	
FT	DISULFID	678	689	
FT	CARBOHYD	248	248	
FT	CARBOHYD	360	360	

FT	CARBOHYD	708	708	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD	1067	1067	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CONFLICT	1025	1025	F -> L (IN AAR53063)			
SO	SEQUENCE	1170 AA:	129646 MM:	0443E493615E7F06 CRC64:			
Query Match		26.6%:	Score 67:	DB 1:			
Best Local Similarity		39.5%:	Pred. No. 1.9:	Length 1170:			
Matches 15:		Conservative 6:	Mismatches 11:	Indels 6:			
Gaps 2:							
QY	13	RCQMOL	-----TSGQMRRC-VSQCDKRFEDIDMSKY 44				
DB	404	RSCDSLNNRCGSSVQTRTCHIQICDCKRFGQDGGMSHW 441					
RESULT	6						
TSPL_XENLA							
ID	TSPL_XENLA	STANDARD:	PRT:	1173 AA.			
AC	P3548;						
DT	01-JUN-1994 (Rel. 29, Created)						
DT	01-JUN-1994 (Rel. 29, Last sequence update)						
DT	01-OCT-1996 (Rel. 34, Last annotation update)						
DE	THROMBOSPONDIN 1 PRECURSOR.						
OS	THBS1 OR TSP1.						
OS	Xenopus laevis (African clawed frog).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;						
OC	Xenopodinae; Xenopus.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	urry L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.:						
RL	Submitted (XXX-1993) to the EMBL/Genbank/DBJ databases.						
CC	-1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND						
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,						
CC	LAMININ AND TYPE V COLLAGEN.						
CC	-1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.						
CC	-1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.						
CC	-1- SIMILARITY: CONTAINS 1 VMPC DOMAIN.						
CC	-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.						
CC	-1- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS						
CC	(WHICH BIND CALCIUM).						
-----							
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>						
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).						
-----							
CC	EMBL: L04278; -; NOT_ANNOTATED_CDS.						
DR	HSP: P00740; .11XA.						
DR	INTERPRO: IPR000561; -						
DR	INTERPRO: IPR000884; -						
DR	INTERPRO: IPR001007; -						
DR	PFAM: PF00008; EGF; 2.						
DR	PFAM: PF00090; tsp_1; 3.						
DR	PFAM: PF00093; vwc_1.						
DR	PROSITE: PS00022; EGF_1; FALSE_NEG.						
DR	PROSITE: PS01186; EGF_2; 1.						
DR	PROSITE: PS50092; TSP1; 3.						
DR	PROSITE: PS01208; VMPC; 1.						
KW	glycosylated; cell adhesion; Calcium-binding; Heparin-binding; Repeat;						
KW	EGF-like domain; Signal.						
FT	SIGNAL	1	22	POTENTIAL.			
FT	CHAIN	23	1173	THROMBOSPONDIN 1.			
FT	DOMAIN	23	235	HEPARIN-BINDING (POTENTIAL).			
FT	DOMAIN	319	376	VMPC.			
FT	DOMAIN	382	546	3 X TSP TYPE-1 REPEATS (CS-LIKE).			
FT	DOMAIN	550	693	3 X EGF-TYPE REPEATS.			
FT	DOMAIN	726	953	7 X TSP TYPE-3 REPEATS (CA-BINDING).			
FT	DOMAIN	954	1173	C-TERMINAL.			
FT							

```

FT REPEAT 438 494 TSP TYPE-1 2.
FT REPEAT 495 546 TSP TYPE-1 3.
FT DOMAIN 550 590 EGF-LIKE 1.
FT DOMAIN 591 648 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 649 693 EGF-LIKE 3.
FT REPEAT 726 761 TSP TYPE-3 1.
FT REPEAT 762 784 TSP TYPE-3 2.
FT REPEAT 785 820 TSP TYPE-3 3.
FT REPEAT 821 843 TSP TYPE-3 4.
FT REPEAT 844 881 TSP TYPE-3 5.
FT REPEAT 882 917 TSP TYPE-3 6.
FT REPEAT 918 953 TSP TYPE-3 7.
FT SITE 929 931 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 554 565 BY SIMILARITY.
FT DISULFID 559 575 BY SIMILARITY.
FT DISULFID 578 589 BY SIMILARITY.
FT DISULFID 595 611 BY SIMILARITY.
FT DISULFID 602 620 BY SIMILARITY.
FT DISULFID 623 647 BY SIMILARITY.
FT DISULFID 653 666 BY SIMILARITY.
FT DISULFID 660 679 BY SIMILARITY.
FT CARBOHYD 681 692 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 705 705 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64;

```

```

Query Match 25.8%; Score 65; DB 1; Length 1173;
Best Local Similarity 39.5%; Pred. No. 3.2;
Matches 15; Conservative 5; Mismatches 12; Indels 6; Gaps 2;

```

```

OY 13 ROCKMOL-----TSGQMRC-VSCDKRFEEDIMSKY 44
      | | | | | | | | | | | | | | | | | |
Db 407 RSCDNLNPGCESSVOTRSCQIDCKRFKODGMSHW 444

```

```

RESULT 7
SBP_SOYBN STANDARD; PRT; 524 AA.
AC 004672;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE SUCROSE-BINDING PROTEIN PRECURSOR (SBP).
GN SBP.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 30-50.
RC TISSUE=EMBRYO;
RX MEDLINE=93104680; Pubmed=1467654;
RA Grimes H.D., Overvoorde P.J., Ripp K., Franceschi V.R., Hitz W.D.;
RT "A 62-kD sucrose binding protein is expressed and localized in
RT tissues actively engaged in sucrose transport.";
RL Plant Cell 4:1561-1574(1992).
CC -1- FUNCTION: PLAYS A ROLE IN SUCROSE TRANSPORT.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: ASSOCIATED WITH THE PLASMA MEMBRANE OF
CC SEVERAL CELL TYPES ENGAGED IN SUCROSE TRANSPORT, INCLUDING THE
CC MESOPHYLL CELLS OF YOUNG SINK LEAVES, THE COMPANION CELLS OF
CC MATURE PHLOEM AND THE CELLS OF DEVELOPING COTYLEDONS.
CC -1- DEVELOPMENTAL STAGE: IN THE COTYLEDON, EXPRESSION IS NOT DETECTED
CC UNTIL 10 DAYS AFTER FERTILIZATION, BETWEEN 10-19 DAYS AFTER
CC FERTILIZATION, EXPRESSION INCREASES RAPIDLY BUT DECLINES 20-30
CC DAYS AFTER FERTILIZATION. 30 DAYS AFTER FERTILIZATION, NO
CC EXPRESSION OCCURS. THIS EXPRESSION PATTERN CLOSELY PARALLELS THE

```

```

CC CC RATE OF SUCROSE UPTAKE IN THE COTYLEDON.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; L06038; AAB03894.1; -.
DR PIR; J01730; J01730.
DR HSSP; P50477; ICAM.
DR INTERPRO; IPR00113; -.
DR Pfam; PF00546; Seedstore_7s; 1.
KW Transport; Sugar transport; Signal; Membrane.
FT SIGNAL 1 29
FT CHAIN 30 524 SUCROSE-BINDING PROTEIN.
SQ SEQUENCE 524 AA; 60522 MW; 0251EE90796EF341 CRC64;

```

```

Query Match 25.0%; Score 63; DB 1; Length 524;
Best Local Similarity 38.2%; Pred. No. 2.6;
Matches 13; Conservative 4; Mismatches 15; Indels 2; Gaps 1;

```

```

OY 2 EFDROEYECRCQMOL--TSGQMRRCVSQCDK 33
      | | | | | | | | | | | | | | | | | |
Db 34 EEDPELVTCMHQCCQDQYTEGDKRVCLQSCDR 67

```

```

RESULT 8
ATIL_VACCV STANDARD; PRT; 724 AA.
AC P24759;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE 94 KDA A-TYPE INCLUSION PROTEIN (ATI).
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91310644; Pubmed=1856205;
RA Amegadzie B.Y., Ahn B.-Y., Moss B.;
RT "Identification, sequence, and expression of the gene encoding a Mr
RT 35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RL J. Biol. Chem. 266:13712-13718(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92124754; Pubmed=1733111;
RA Amegadzie B.Y., Sisler J.R., Moss B.;
RT "Frame-shift mutations within the vaccinia virus A-type inclusion
RT protein gene.";
RL Virology 186:777-782(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92074241; Pubmed=1962448;
RA de Carlos A., Pez E.;
RT "Isolation and characterization of mutants of vaccinia virus with a
RT modified 94-kDa inclusion protein.";
RL Virology 185:768-778(1991).
CC -1- FUNCTION: NATURE VIRUSES ARE OCCURED INTO THE ATI, AND IT HAS
CC BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
CC DISSEMINATION FROM ANIMAL TO ANIMAL.
CC -1- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF
CC LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF
CC INFECTED CELLS.
CC -1- SIMILARITY: 92% IDENTITY TO COMPOX VIRUS A-TYPE INCLUSION
CC PROTEIN N-TERMINAL.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M6187; AAA48321.1; -  
DR EMBL: M76371; AAA48275.1; -  
DR EMBL: X57318; CAA40574.1; -  
DR PIR: A41701; MWV294.  
DR PIR: S29908; S29908.  
FT CONFLICT 587 588 KO -> SK (IN REF. 3).  
FT CONFLICT 610 610 R -> H (IN REF. 3).  
FT CONFLICT 618 618 R -> RR (IN REF. 3).  
FT CONFLICT 682 682 S -> R (IN REF. 3).  
SO SEQUENCE 724 AA; 84235 MW; C560CF61C9903028 CRC64;

Query Match 22.6%; Score 57; DB 1; Length 724;  
Best Local Similarity 40.6%; Pred. No. 17;  
Matches 13; Conservative 11; Mismatches 6; Indels 2; Gaps 2;

OY 12 KRCOMLETSQMRRCVSCCKRFEEDIDMSK 43  
DB 550 KRRNWEWELS-RLRRDIKECDK-YKEDLDKAK 579

## RESULT 9

ATI\_CAMPC STANDARD; PRT; 726 AA.  
ID ATI\_CAMPC  
AC 005482;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE A-TYPE INCLUSION PROTEIN (ATI).  
OS Camelox virus (strain Cp-1).  
OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Orthopoxvirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93346980; Pubmed=8345359;  
RA Meyer H., Rziha H.-J.;  
RT "Characterization of the gene encoding the A-type inclusion protein  
RT of camelox virus and sequence comparison with other  
RT orthopoxviruses."  
RL J. Gen. Virol. 74:1679-1684(1993).  
CC -1- FUNCTION: MATURE VIRUSES ARE OCCLUDED INTO THE ATI, AND IT HAS  
CC BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING  
CC DISSEMINATION FROM ANIMAL TO ANIMAL.  
CC -1- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF  
CC LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF  
CC INFECTED CELLS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 22.6%; Score 57; DB 1; Length 726;  
Best Local Similarity 40.6%; Pred. No. 17;  
Matches 13; Conservative 11; Mismatches 6; Indels 2; Gaps 2;

OY 12 KRCOMLETSQMRRCVSCCKRFEEDIDMSK 43  
DB 551 KRRNWEWELS-RLRRDIKECDK-YKEDLDKAK 580

## RESULT 10

ATI\_COMPCX STANDARD; PRT; 1284 AA.  
ID ATI\_COMPCX  
AC P16602;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE A-TYPE INCLUSION PROTEIN (ATI).  
CN ATI.  
OS Cowpox virus (CPV).  
OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Orthopoxvirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CEP05;  
RX MEDLINE=88089536; Pubmed=2826668;  
RA Funahashi S., Sato T., Shida H.;  
RT "Cloning and characterization of the gene encoding the major protein  
RT of the A-type inclusion body of cowpox virus."  
RL J. Gen. Virol. 69:35-47(1988).  
RN [2]  
RP SEQUENCE OF 1-109 FROM N.A.  
RX MEDLINE=88111568; Pubmed=2828037;  
RA Patel D.D., Pickup D.J.;  
RT "Messenger RNAs of a strongly-expressed late gene of cowpox virus  
RT contain 5'-terminal poly(A) sequences."  
RL EMBO J. 6:3787-3794(1987).  
CC -1- FUNCTION: MATURE VIRUSES ARE OCCLUDED INTO THE ATI, AND IT HAS  
CC BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING  
CC DISSEMINATION FROM ANIMAL TO ANIMAL.  
CC -1- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF  
CC LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF  
CC INFECTED CELLS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 22.6%; Score 57; DB 1; Length 1284;  
Best Local Similarity 40.6%; Pred. No. 29;  
Matches 13; Conservative 11; Mismatches 6; Indels 2; Gaps 2;

OY 12 KRCOMLETSCOMRCVSOCCKREEDIDMSK 43  
 DB 552 KRRNWEELS-RLRRDIECKD-KYEDIDKAK 581

## RESULT 11

PP6C\_RAT STANDARD; PRT; 305 AA.  
 AC 064620;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-1998 (Rel. 36, Last annotation update)  
 DE SERINE/THREONINE PROTEIN PHOSPHATASE 6 (EC 3.1.3.16) (PP6) (PROTEIN PHOSPHATASE V) (PP-V).  
 GN PPP6C OR PPV.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=TESTIS, AND BRAIN;  
 RX MEDLINE=94357899; PubMed=8077208;  
 RA Becker W., Kentrup H., Klump S., Schultz J.E., Joost H.G.:  
 RT "Molecular cloning of a protein serine/threonine phosphatase containing a putative regulatory tetrapeptide repeat domain.";  
 RL J. Biol. Chem. 269:22586-22592(1994).  
 CC -1- FUNCTION: MAY FUNCTION IN CELL CYCLE REGULATION.  
 CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN + ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN SPLEEN, BRAIN AND LUNG.  
 CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-V SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X77236; CAA54453.1; -;  
 DR HSSP: P08129; IFM.  
 DR INTERPRO: IPR000934; -;  
 DR PRAM: PF00149; STphosphatase; 1.  
 DR PRINTS: PR00114; STPHPTASE.  
 DR PROSITE: PS00125; SER\_THR\_PHOSPHATASE; 1.  
 KM Hydrolase; Iron; Manganese.  
 FT METAL 53 53 IRON (BY SIMILARITY).  
 FT METAL 55 55 IRON (BY SIMILARITY).  
 FT METAL 81 81 IRON AND MANGANESE (BY SIMILARITY).  
 FT METAL 113 113 MANGANESE (BY SIMILARITY).  
 FT ACT\_SITE 114 114 GENERAL ACID (BY SIMILARITY).  
 FT METAL 163 163 MANGANESE (BY SIMILARITY).  
 FT METAL 237 237 MANGANESE (BY SIMILARITY).  
 FT METAL 237 237 MANGANESE (BY SIMILARITY).  
 SQ SEQUENCE 305 AA; 35106 MW; A91F9939FDBDF110 CRC64;

Query Match 21.8%; Score 55; DB 1; Length 305;  
 Best Local Similarity 38.2%; Pred. No. 13;  
 Matches 13; Conservative 2; Mismatches 19; Indels 0; Gaps 0;

OY 4 DROEYECRCOMLETSCOMRCVSOCCKRFEF 37  
 DB 5 DLDKYVEIAROCKYLPENDLKLCDYVCDLLEE 38

RESULT 12  
 ENPL\_CHICK STANDARD; PRT; 795 AA.  
 ID ENPL\_CHICK

AC P08110; Q90870; Q90869;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE ENDOPLASMIN PRECURSOR (HEAT SHOCK 108 KDA PROTEIN) (HSP108) (HSP 108) (TRANSFERIN-BINDING PROTEIN).  
 GN TRAI.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauilia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87076542; PubMed=3024703;  
 RA Kilomaa M.S., Weigel N.L., Kleinsek D.A., Beattie W.G., Conneely O.M., March C., Zarucki-Schulz T., Schnader W.T., O'Malley B.W.:  
 RT "Amino acid sequence of a chicken heat shock protein derived from the complementary DNA nucleotide sequence.";  
 RL Biochemistry 25:6244-6251(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=94220155; PubMed=8166742.  
 RA Hayes G.R., Himpler B.S., Weiner K.X.B., Lucas J.J.:  
 RT "A chicken transferrin binding protein is heat shock protein 108.";  
 RL Biochem. Biophys. Res. Commun. 200:65-70(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVIDUCT;  
 RX MEDLINE=87117509; PubMed=3027654;  
 RA Kleinsek D.A., Beattie W.G., Tsai M.J., O'Malley B.W.:  
 RT "Molecular cloning of a steroid-regulated 108k heat shock protein gene from hen oviduct.";  
 RL Nucleic Acids Res. 14:10053-10069(1986).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVIDUCT;  
 RA Forsgren M.:  
 RL Submitted (XXX-1987) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M14772; AAA48826.1; -;  
 DR EMBL: M31321; AAA48827.1; -;  
 DR EMBL: X04961; CAA28629.1; -;  
 DR PIR: A24461; HHC08.  
 DR PIR: J02205; J02205.  
 DR HSSP: P02829; JAH8.  
 DR INTERPRO: IPR000886; -;  
 DR INTERPRO: IPR001404; -;  
 DR PRAM: PF00183; HSP90; 1.  
 DR PRINTS: PR00775; HEATSHOCK90.  
 DR PROSITE: PS00014; ER\_TARGET; 1.  
 DR PROSITE: PS00298; HSP90; 1.  
 KM Chaperone; Endoplasmic reticulum; Glycoprotein; Calcium binding; K<sup>+</sup> signal; Heat shock.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 795 ENDOPLASMIN.  
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 792 795 PREVENT SECRETION FROM ER.

```

FT CONFLICT 206 206 D -> E (IN REF. 3 AND 4).
FT CONFLICT 267 267 V -> L (IN REF. 3).
FT CONFLICT 303 303 E -> Q (IN REF. 3).
FT CONFLICT 307 307 N -> D (IN REF. 3 AND 4).
FT CONFLICT 317 317 E -> H (IN REF. 3).
FT CONFLICT 378 378 G -> A (IN REF. 3).
FT CONFLICT 593 594 EG -> DR (IN REF. 3).
FT CONFLICT 653 653 W -> C (IN REF. 3).
FT CONFLICT 669 675 GNDISTN -> VFSS (IN REF. 4).
SQ SEQUENCE 795 AA: 91555 MW: BEL1B29E1DBEC5A9A CRC64;

```

```

Query Match 21.8%; Score 55; DB 1; Length 795;
Best Local Similarity 23.8%; Pred. No. 32;
Matches 10; Conservative 12; Mismatches 20; Indels 0; Gaps 0;

```

```

OY 2 EFDROEYECCKROCMOLETSGOMRCVSCQDKRFEDIDMSK 43
DB 581 EFDGKRFQNVAKGCVKFESEKSKSEKREALEKEFEPLIMMK 622

```

```

RESULT 13
YL53.CAEEL STANDARD; PRT; 244 AA.
AC P34453;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 29.0 KDA PROTEIN F44E2.3 IN CHROMOSOME III.
GN F44E2.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smalldon N., Smith A., Smith M., Soudhamer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaughan M.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO DNAB.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: L23646; AAA28041.1; -
DR PIR: S44822; S44822.
DR MORPEP: F44E2.3; CE00181.
KW Hypothetical protein.
FT DOMAIN 3 45 ARG/ASP/LYS-RICH.
FT 79 90 PRO-RICH.
SQ SEQUENCE 244 AA: 28994 MW: C0CE677FB01A2B18 CRC64;

```

```

Query Match 21.6%; Score 54.5; DB 1; Length 244;
Best Local Similarity 39.4%; Pred. No. 12;
Matches 13; Conservative 8; Mismatches 7; Indels 5; Gaps 2;

```

```

OY 9 ECKRCOCMOLETSQOMRCVSCQDKRFEDIDMSK 38
DB 164 EESRKKCKOLE-ALERKVLVEAESRKKFEED 194

```

```

RESULT 14
YL68.CAEEL STANDARD; PRT; 1021 AA.
AC P46582;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 111.9 KDA PROTEIN C34E10.8 IN CHROMOSOME III.
GN C34E10.8
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Kirsten J.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U10402; AAA19070.1; -
CC MORPEP: C34E10.8; CE01188.
CC KW Hypothetical protein.
SQ SEQUENCE 1021 AA: 111858 MW: F1B63DAA26332F5F CRC64;

```

```

Query Match 21.6%; Score 54.5; DB 1; Length 1021;
Best Local Similarity 22.0%; Pred. No. 46;
Matches 11; Conservative 15; Mismatches 13; Indels 11; Gaps 2;

```

```

OY 1 SEFDROEYECCKR-----QCMOLETSGOMRCVSCQDKRFEDIDMSK 43
DB 24 NEYDKNKRYGOCNAYSYRTLVRCQIRSKELARHGRC-----BEHVEFSK 69

```

```

RESULT 15
CGDL_BRARE STANDARD; PRT; 291 AA.
AC C0D1_BRARE
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE G1/S-SPECIFIC CYCLIN D1.
GN CYCDL.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96138542; PubMed-8547308;
RA Yarden A., Salomon D., Geiger B.;
RT "zebrafish cyclin D1 is differentially expressed during early
RT embryogenesis."
RL Biochim. Biophys. Acta 1264:257-260(1995).
CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
CC (START) TRANSITION.
CC -1- SUBUNIT: INTERACTS WITH THE CDK4 AND CDK6 PROTEIN KINASES TO FORM
CC A SERINE/THREONINE KINASE HOLOENZYME COMPLEX. THE CYCLIN SUBUNIT
CC IMPARTS SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN D SUBFAMILY.

```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
DR EMBL; X87581; CAA60885.1; -;  
DR ZFIN; ZDB-GENE-980526-176; CYCD1.  
DR INTERPRO; IPR000553; -;  
DR PFAM; PF00134; CYCLIN; 1.  
DR PROSITE; PS00292; CYCLINS; 1.  
KW Cyclic; Cell cycle; Cell division.  
SQ SEQUENCE 291 AA; 33067 MW; FA5274CB1B46D5EF CMC64;

Query Match	21.48;	Score 54;	DB 1;	Length 291;
Best Local Similarity	38.50;	Pred. NO. 16;		
Matches 15; Conservative	7;	Mismatches 11;	Indels 6;	Gaps 2;

```

OY      10  ECKRQCMQ-----LETS-GQMRRCVSCQDKRFEEIDWS 42
        : | | : ||| | : : : | | ||| : |
Db      242 DCLRSCEQIESLLESSLRQQAQHISTETKRVEEVDLS 280

```

```
Search completed: March 1, 2001, 16:16:27
Job time: 407 sec
```

